Convolution particle filters for parameter estimation in general state-space models

Fabien Campillo, Vivien Rossi

PUBLICATION
INTERNE
N° 1808

CONVOLUTION PARTICLE FILTERS FOR PARAMETER ESTIMATION IN GENERAL STATE-SPACE MODELS

FABIEN CAMPILLO AND VIVIEN ROSSI

ISSN 1166-8687

IRISA
CAMPUS UNIVERSITAIRE DE BEAULIEU - 35042 RENNES CEDEX - FRANCE

INSTITUT DE RECHERCHE EN INFORMATIQUE ET SYSTÈMES ALÉATOIRES
Campus de Beaulieu – 35042 Rennes Cedex – France
Tél. : (33) 02 99 84 71 00 – Fax : (33) 02 99 84 71 71
http://www.irisa.fr

Convolution particle filters for parameter estimation in general state-space models

Fabien Campillo * and Vivien Rossi **

Systèmes numériques
Projets Aspi

Publication interne n ° 1808 — Juin 2006 — 25 pages

Abstract: The state-space modeling of partially observed dynamic systems generally requires estimates of unknown parameters. From a practical point of view, it is relevant in such filtering contexts to simultaneously estimate the unknown states and parameters.

Efficient simulation-based methods using convolution particle filters are proposed. The regularization properties of these filters is well suited, given the context of parameter estimation. Firstly the usual non Bayesian statistical estimates are considered: the conditional least squares estimate (CLSE) and the maximum likelihood estimate (MLE). Secondly, in a Bayesian context, a Monte Carlo type method is presented. Finally these methods are compared in several simulated case studies.

Key-words: Hidden Markov models, parameter estimation, particle filter, convolution kernels, conditional least squares estimate, maximum likelihood estimate

(Résumé : tsvp)

* Fabien.Campillo@inria.fr
** Vivien.Rossi@irisa.fr

Centre National de la Recherche Scientifique
(UMR 6074) Université de Rennes 1 – Insa de Rennes

Institut National de Recherche en Informatique
et en Automatique – unité de recherche de Rennes

Filtres particulaires à convolution pour l'estimation de paramètres dans des modèles à espace d'état généraux

Résumé : La modélisation par espace d'état de systèmes dynamiques partiellement observés requière le plus souvent l'estimation de paramètres inconnus. En pratique, il est pertinent dans un tel cadre de simultanément estimer l'état non observé et les paramètres inconnus.

On propose des méthodes de simulation faisant appel à des filtres particulaires à convolution. Les propriétés de régularisation de ces filtres sont particulièrement adaptées à ce contexte d'estimation paramétrique. Dans un premier temps, on considère les estimées des moindres carrés conditionnelles et du maximum de vraisemblance. Puis, dans un contexte bayésien, on propose une méthode de type Monte Carlo. Ces méthodes sont enfin comparées sur plusieurs exemples simulés.

Mots clés : modèles de Markov cachés, estimation paramétrique, filtre particulaire, noyaux de convolution, estimation des moindres carrés conditionnel, estimation du maximum de vraisemblance

Contents

1 Introduction

Consider a general state-space dynamical system described by an unobserved state process x_t and an observation process y_t taking values in $\mathbb{R}^d$ and $\mathbb{R}^q$ respectively. This system depends on an unknown parameter $\theta \in \mathbb{R}^p$. Suppose that the state process is Markovian, and that the observations y_t are independent conditionally to the state process. Suppose also that the distribution law of y_t depends only on x_t. Hence this system is completely described by the state process transition density and the emission density, namely

$$\begin{array}{rcl} x_t|x_{t-1} & \sim & f_t(x_t|x_{t-1},\theta)\,, \\ y_t|x_t & \sim & h_t(y_t|x_t,\theta)\,, \end{array} \tag{1}$$

and by the initial density law π_0 of x_0.

The goal is to estimate simultaneously the parameter θ and the state process x_t based on the observations $y_{1:t} = \{y_1, \ldots, y_t\}$.

In the nonlinear hidden processes framework, the parameter estimation procedure is often based on an approximation of the optimal filter. The extended Kalman filter and its various alternatives can give good results in practice but suffer from an absence of theoretical backing. The particle filters propose a good alternative: in many practical cases they give better results, moreover their theoretical properties are becoming increasingly well understood [1] [2] [3].

It is thus particularly appealing to use particle filtering in order to estimate parameters in partially observed systems. For a review of the question, one can consult Doucet [4] or Liu & West [5]. There are two main approaches:

- The non Bayesian approach which consists of minimizing a given cost function like the conditional least squares criterion or by maximizing the likelihood function. These methods are usually performed in batch processes but can also be extended to recursive procedures.

- The Bayesian approach where an augmented state variable which includes the parameter is processed by a filtering procedure. These methods suppose that a prior law is given for the parameter and are performed on-line.

In practice, the first approach could be used as an initialization for the second one.

Due to the partially observed system framework, the objective function introduced in the first approach should be approximated for various values of the parameter θ. This is done via the particle approximation of the conditional law $p(y_t|y_{1:t-1},\theta)$. The Monte Carlo nature of this particle approximation will make the optimization problematic. However, recent analyses propose significant improvements of these aspects [6] [4].

The second approach takes place in a classical Bayesian framework, a prior probability law $\rho(\theta)$ is thus introduced on the parameter θ. A new state variable (x_t,θ_t), joining all the unknown quantities, is considered and the posterior law $p(x_t,\theta_t|y_{1:t})$ is then approximated using particle filters.

In this paper we propose and compare different estimates corresponding to these two approaches and based on convolution particle filter introduced in [7].

The paper is organized as follows, we first recall the principle of the convolution filter for the dynamical systems without unknown parameters. The application and the convergence analysis of this filter require weaker assumptions than the usual particle filters. This is due to the use of convolution kernels to weight the particle.

Then the conditional least squares estimate and the maximum likelihood estimate are presented. Their adaptation to the state-space model context is possible thanks to the convolution filters. The computation of these two estimates calls upon an optimization procedure, which is problematic for particle filters because of their random nature.

Next, the Bayesian estimation approach is presented, it also relies on the convolution particle filter. In this context, the standard particle approach highlights various drawbacks that are avoided by the convolution thanks to their smooth nature.

Finally, these various estimates are compared in a range of simulated cases.

2 The convolution filters

To present the convolution filter, suppose that the parameter θ is known and consider:

$$\begin{array}{rcl} x_t|x_{t-1} & \sim & f_t(x_t|x_{t-1})\,, \\ y_t|x_t & \sim & h_t(y_t|x_t)\,. \end{array} \tag{2}$$

The objective is to estimate recursively the optimal filter

$$p(x_t|y_{1:t}) = \frac{p(x_t, y_{1:t})}{p(y_{1:t})} = \frac{p(x_t, y_{1:t})}{\int p(x_t, y_{1:t})\,dx_t} \tag{3}$$

where $p(x_t, y_{1:t})$ is the $(x_t, y_{1:t})$ joint density.

Assumption: Suppose that we know how to sample from the laws $f_t(\cdot|x_{t-1})$, $h_t(\cdot|x_t)$ and also from the initial law π_0.

Note that the explicit description of the conditional densities f_t and h_t is useless whereas for the standard particle filtering approaches h_t should be stated explicitly. For example, in case of observation equations like $y_t = H(x_t, v_t)$ or $H(x_t, y_t, v_t) = 0$, where v_t is a noise, the conditional density h_t is in general not available.

2.1 The simple convolution filter (CF)

Let $\{x_0^i\}_{i=1\cdots n}$ be a sample of size n of π_0. For all $i = 1\cdots n$, starting from x_0^i, t successive simulations from the system (2) lead to a sample $\{x_t^i, y_{1:t}^i\}_{i=1\cdots n}$ from $p(x_t, y_{1:t})$. We get the following empirical estimate of the joint density:

$$p(x_t, y_{1:t}) \simeq \frac{1}{n}\sum_{i=1}^{n} \delta_{(x_t^i, y_{1:t}^i)}(x_t, y_{1:t}) \tag{4}$$

for $t = 0$
  initial sampling: $x_0^1 \cdots x_0^n \sim \pi_0$
  weight initialization: $w_0^i \leftarrow 1$ for $i = 1 : n$

for $t \geq 1$
  for $i = 1 : N$
    state sampling: $x_t^i \sim f_t(\cdot|x_{t-1}^i)$
    observation sampling: $y_t^i \sim h_t(\cdot|x_t^i)$
    weight updating: $w_t^i \leftarrow w_{t-1}^i \, K_{h_n}^y(y_t - y_t^i)$

  filter updating:
$$p_t^n(x_t|y_{1:t}) = \frac{\sum_{i=1}^n w_t^i \, K_{h_n}^x(x_t - x_t^i)}{\sum_{i=1}^n w_t^i}$$

Table 1: The simple convolution filter (CF).

where δ_x is the Dirac measure in x.

The Kernel estimate $p_t^n(x_t, y_{1:t})$ of $p(x_t, y_{1:t})$ is then obtained by convolution of the empirical measure (4) with an appropriate kernel (cf. Appendix A):

$$p_t^n(x_t, y_{1:t}) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n K_{h_n}^x(x_t - x_t^i) \; K_{h_n}^{\bar{y}}(y_{1:t} - y_{1:t}^i)$$

where
$$K_{h_n}^{\bar{y}}(y_{1:t} - y_{1:t}^i) \stackrel{\text{def}}{=} \prod_{s=1}^t K_{h_n}^y(y_s - y_s^i) \, .$$

in which $K_{h_n}^x$, $K_{h_n}^y$ are Parzen-Rosenblatt kernels of appropriate dimensions. Note that in $K_{h_n}^x(x_t - x_t^i)$ (resp. $K_{h_n}^y(y_t - y_t^i)$) h_n could implicitly depend on n, d and $x_t^{1:n}$ (resp. n, q and $y_t^{1:n}$) (see Section 2.3).

From (3), an estimate of the optimal filter is then:

$$p_t^n(x_t|y_{1:t}) \stackrel{\text{def}}{=} \frac{\sum_{i=1}^n K_{h_n}^x(x_t - x_t^i) \; K_{h_n}^{\bar{y}}(y_{1:t} - y_{1:t}^i)}{\sum_{i=1}^n K_{h_n}^{\bar{y}}(y_{1:t} - y_{1:t}^i)} \tag{5}$$

The basic convolution filter (CF) is defined by the density estimate (5). A simple recursive algorithm for its practical computation is presented in Table 1.

Convergence properties of $p_t^n(x_t|y_{1:t})$ to the optimal filter are ensured [7] when $h_n \to 0$ and $n h_n^{tq+d} \to \infty$. Just like the Monte Carlo filters without resampling, it implies that n must grow with t to maintain a good estimation. A better approach with a resampling step is proposed in the next section.

for $t = 0$
filter initialization: $p_0^n \leftarrow \pi_0$

for $t \geq 1$
resampling: $\bar{x}_{t-1}^1 \cdots \bar{x}_{t-1}^n \sim p_{t-1}^n$
state sampling: $x_t^i \sim f_t(\cdot|\bar{x}_{t-1}^i)$ for $i = 1:n$
observation sampling: $y_t^i \sim h_t(\cdot|x_t^i)$ for $i = 1:n$
filter updating:

$$p_t^n(x_t|y_{1:t}) = \frac{\sum_{i=1}^n K_{h_n}^y(y_t - y_t^i)\; K_{h_n}^x(x_t - x_t^i)}{\sum_{i=1}^n K_{h_n}^y(y_t - y_t^i)}$$

Table 2: The resampled convolution filter (R-CF).

2.2 The resampled convolution filter (R-CF)

A resampling step can take place very easily at the beginning of each time step of the basic CF algorithm, the resulting procedure is presented in Table 2

Convergence properties to the optimal filter are also established in [7]. It is necessary that $h_n^{2q} = O(n^{-\alpha})$, with $\alpha \in (0,1)$ and $nh_n^{q+d}/\log n \to \infty$ to ensure the L_1-convergence of the R-CF to the optimal filter (here the bandwidth parameters are the same for the state and the observation spaces).

2.3 Comments

The practical use of the CF and R-CF filters requires the choice of the kernel functions K^x, K^y and of the bandwidth parameters h_n^x, h_n^y. The nature of the kernel does not appreciably affect the quality of the results.

The choice $h_n^x = C_x \times n^{-1/(4+d)}$, $h_n^y = C_y \times n^{-1/(4+q)}$ is optimal for the mean square error criterion. The choice of the C's is a critical issue for density estimation and sophisticated techniques have been proposed (see, e.g., [8]). In the on-line context of nonlinear filtering these techniques are not usable. Moreover, particle filtering is aimed to “track” the state and not really to sharply estimate the conditional density.

The generic form $C_x = c_x \times [\mathrm{Cov}(x_t^1, \ldots, x_t^n)]^{1/2}$, $C_y = c_y \times [\mathrm{Cov}(y_t^1, \ldots, y_t^n)]^{1/2}$ with c_x, $c_y \simeq 1$ gives good results. For the simulations of the last section, we take a Gaussian kernel and we will see that the c's are easily adjusted.

3 Conditional least squares estimate

The standard least squares estimate is not obtainable here since only the y_t's are available and, moreover, they are dependent. Thus let us consider a conditional least squares estimate, introduced to treat the time series (see Tong [9]).

3.1 The theoretical estimate

Let $\{y_t\}_{t\geq 1}$ the stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P}_\theta)$, whose distribution depends on the parameter $\theta \in \mathbb{R}^p$.

Let θ_* the true value of the parameter. The conditional least squares estimate of θ is the value $\hat{\theta}_T$ which minimizes

$$Q_T(\theta) \stackrel{\text{def}}{=} \sum_{t=1}^{T} |y_t - \hat{y}_t(\theta)|^2 \tag{6}$$

where

$$\hat{y}_t(\theta) \stackrel{\text{def}}{=} \mathbb{E}_\theta[y_t|y_{1:t-1}]$$

with $\mathbb{E}_\theta[y_1|y_{1:0}] = \mathbb{E}_\theta[y_1]$. The demonstration of the convergence of $\hat{\theta}_T$ to θ_* under $\mathbb{P}_{\theta_*}$, as $T \to \infty$ is detailed in [9] and [10]. In general, and especially in our context, the quantity $\mathbb{E}_\theta[y_t|y_{1:t-1}]$ is unreachable. It can be estimated using a particle filter. Such an estimate based on the CF is built in the following section.

3.2 The practical estimate

The conditional density of y_t given $y_{1:t-1}$ is

$$p(y_t|y_{1:t-1},\theta) = \frac{p(y_{1:t}|\theta)}{p(y_{1:t-1}|\theta)} = \frac{p(y_{1:t}|\theta)}{\int p(y_{1:t}|\theta)\,dy_t}$$

so that

$$\hat{y}_t(\theta) = \frac{\int y_t\, p(y_{1:t}|\theta)\,dy_t}{\int p(y_{1:t}|\theta)\,dy_t} \tag{7}$$

For θ and $t \geq 1$ given, it is possible to generate n trajectories $(x^i_{0:t}, y^i_{1:t})$, for $i = 1 \cdots n$, according to (1). Finally $\frac{1}{n}\sum_{i=1}^n y^i_t\, K^{\bar{y}}_{h_n}(y_{1:t-1} - y^i_{1:t-1})$ and $\frac{1}{n}\sum_{i=1}^n K^{\bar{y}}_{h_n}(y_{1:t-1} - y^i_{1:t-1})$ are respectively the convolution kernel estimates of the numerator and denominator in (7). Hence the estimate of $\hat{y}_t(\theta)$ built from these n trajectories is

$$\hat{y}^n_t(\theta) \stackrel{\text{def}}{=} \frac{\sum_{i=1}^n y^i_{t+1}\, K^{\bar{y}}_{h_n}(y_{1:t-1} - y^i_{1:t-1})}{\sum_{i=1}^n K^{\bar{y}}_{h_n}(y_{1:t-1} - y^i_{1:t-1})}\,.$$

We take

$$\hat{Q}^n_T(\theta) \stackrel{\text{def}}{=} \sum_{t=1}^{T} |y_t - \hat{y}^n_t(\theta)|^2 \tag{8}$$

to estimate the function Q_T. The associated least squares estimate is then $\hat{\theta}^n_T = \arg\min_\theta \hat{Q}^n_T(\theta)$.

To obtain results of convergence, it is necessary to introduce an assumption like the uniform convergence of $\hat{Q}_T^n$ to Q_T, because we need the convergence of the "argmin". Such an assumption is difficult to check in practice since it depends primarily on the dynamic system considered and of the role held by the parameter. A thorough study of the theoretical properties of the estimate is done in [11].

4 Maximum likelihood estimate

The likelihood function is by definition:

$$L_T(\theta) \stackrel{\text{def}}{=} p(y_{1:T}|\theta) = p(y_1|\theta) \prod_{t=2}^{T} p(y_t|y_{1:t-1}, \theta)\,. \tag{9}$$

Of course this function is not generally computable, it is then necessary to have recourse to estimation, see Kitagawa [12] [13].

The practical likelihood estimate depends on the type of convolution filter used, this is detailed in the following sections.

4.1 Maximum likelihood estimation with the CF

In the CF case an immediate estimate is:

$$\hat{L}_T^n(\theta) \stackrel{\text{def}}{=} p^n(y_{1:T}|\theta) = \frac{1}{n} \sum_{i=1}^{n} K_{h_n}^{\bar{y}}(y_{1:T} - y_{1:T}^i)$$

Thus $\hat{\theta}_T^n = \arg\max_\theta \hat{L}_T^n(\theta)$ approximates the maximum likelihood estimate.

4.2 Maximum likelihood estimation with the R-CF

For the R-CF formalization is not immediate. However all the quantities necessary to compute an estimate are made available with the R-CF algorithm. Indeed, the variables $\{y_{t+1}^i\}_{i=1\cdots n}$ generated in the observation sampling step of the R-CF algorithm are realizations of $p^n(y_{t+1}|y_{1:t}, \theta)$. Thus by applying a convolution kernel to $\{y_{t+1}^i\}_{i=1\cdots n}$, we obtain the following estimate of the likelihood function:

$$\hat{L}_T^{n,r}(\theta) = \prod_{t=1}^{T} \frac{1}{n} \sum_{i=1}^{n} K_{h_n}^{y}(y_t - y_t^i) \tag{10}$$

and $\hat{\theta}_T^{n,r} = \arg\max_\theta \hat{L}_T^{n,r}(\theta)$ approximates the maximum likelihood estimate.

As for the conditional least squares estimate, it is necessary to introduce an assumption of uniform convergence in θ to ensure the convergence of $\hat{\theta}_T^n$ and $\hat{\theta}_T^{n,r}$. For more details see [11].

5 Optimization difficulties

For each fixed value of θ, approximations (8) and (10) of the least squares function (6) and of the likelihood function (9) are computed through kernel particle filters with sampling procedures based on laws depending on θ. These approximations will not be as smooth as their original counterparts and standard optimization procedures will severely fail in such a context. Therefore, it is necessary to use specific optimization techniques.

This issue can be addressed by stochastic approximation and optimization methods. Recently Doucet [4] proposed a Robbins-Monroe procedure in this HMM framework. The principal defect of these approaches is the slowness of their convergence, in spite of the efforts to improve this aspect, the computing times remain high in practice.

When the random quantities in the dynamic system, generally the noises, are independent of the other quantities, it is possible to freeze their values to one of their realizations so that the functions to optimize in θ is not stochastic any more. This technique can only be applied to CF filter, indeed for the R-CF it is impossible to freeze the resampling steps. Hence, because of the particle impoverishment of the CF filter described above, this algorithm is only valid for short length time series.

This approach is connected with techniques of optimization on MCMC estimates. The principle is as follows: for every time t, the simulated random quantities are frozen to their realizations, it is then possible to use the traditional minimization algorithms like Gauss-Newton. The parameter estimation is thus obtained for a given random realization. The study of this method for static optimization is carried out in Geyer [14].

An adaption to the sequential context of nonlinear filtering, for the maximization of the likelihood, is proposed by Hürzeler & Künsch [15]. Several problems arise in practice, for example, for some values of the parameters, all the particle weights can be low providing a poor quality estimate. This approach remains extremely attractive as it then becomes possible to carry out optimizations using only one sample and consequently is valid for small variations of the parameter value. Thus Cérou et al [6] proposed an estimate of the derived filter based on this principle.

Of course, it is also possible to use a stochastic version of EM algorithm for this type of problem of optimization. Some references for this alternative are proposed in [15], but the difficulty of implementation makes it unfeasible.

6 R-CF with unknown parameters approach

Suppose that the parameter θ is a random variable with a given prior law $\rho(\theta)$ and consider the augmented state variable (x_t, θ_t) with the following dynamic:

$$\theta_t = \theta_{t-1}\,, \qquad \theta_0 \sim \rho\,, \tag{11a}$$
$$x_t|x_{t-1} \sim f_t(x_t|x_{t-1}, \theta_t)\,, \qquad x_0 \sim p_0\,, \tag{11b}$$
$$y_t|x_t \sim h_t(y_t|x_t, \theta_t)\,. \tag{11c}$$

The posterior law of θ_t is then given by the nonlinear filter.

The constant dynamic (11a) may lead to the divergence of the standard particle filters. This is due to the fact that the parameter space is only explored at the initialization step of the particle filter which causes the impoverishment of the variety of the relevant particles. Among the approaches proposed to avoid this trouble, Storvik [16] marginalizes parameters out of the posterior distribution then assume that the concerned parameters depend on sufficient statistics which allows their simulations and avoids the impoverishment. However it is not practically useful for general systems. Kitagawa [12] and Higuchi [17] set an artificial dynamic on the parameter, like $\theta_t = \theta_{t-1} + \zeta_t$ or more complex, thus risking mismatching the system dynamic. Gilks & Berzuini [18], Lee & Chia [19] add a Markov chain Monte Carlo procedure to increase the particle diversity, but this is cumbersome. To avoid these additions West [20], Liu & West [5] propose to smooth the empirical measure of the parameter posterior law with a Gaussian distribution.

More generally, regularization techniques are used to avoid the degeneration of the particle filters. Most of the time the regularization only concerns the state variables, see [21] and [22]. However this approach still suffers from some of the restrictions of the traditional methods: it requires the non nullity of the noise variances and the analytical availability of the likelihood function. These restrictions were dropped in [23] by the regularization of the observation model. However, as the state model is not regularized, the approach remains sensitive to the problem of degeneration of the particle filters.

In order to circumvent these two problems simultaneously, Rossi & Vila [7] jointly regularized the state model and the observation model. Their approach can be interpreted as a generalization of the preceding models, thanks to an extension of the concept of particle which includes the state and the observation. However, the construction and the theoretical study of the corresponding filters are different as they are based on the nonparametric estimate of the conditional densities by convolution kernels. The filter used in this section to estimate simultaneously the state and the parameters in (11), extends the results of [7]. It is not necessary for the kernel to be Gaussian as in West [20], any kernel satisfying the conditions of the Appendix A will be valid.

The regularization with convolution kernels can also be viewed as artificial noise. Thus our approach is connected to the methods[12], [17] presented previously. However contrary to these methods, it respects dynamics (11a) and allows convergence results. In terms of artificial noise on dynamics, this means that we have identified a whole family of acceptable noises and that we have also characterized the way in which their variance must decrease to zero.

6.1 Algorithm

The R-CF filter (Table 2) applied to the system (11) leads to the algorithm presented in Table 3. It provides consistent estimates of $p(x_t, \theta_t | y_{1:t})$, $p(x_t | y_{1:t})$ and $p(\theta_t | y_{1:t})$. The first probability law is the key element of the algorithm. It is used as a sample generator and it is updated at every time iteration. The two last laws are used to estimate the state x_t and the parameter θ_t respectively.

Generate $\bar{x}_0^i \sim p(x_0)$ and $\bar{\theta}_0^i \sim \rho(\theta)$ for $i = 1 \cdots n$

For $t = 1$

generation of the trajectories: for $i = 1 \cdots n$

$x_1^i \sim f_1(\cdot|\bar{x}_0^i, \bar{\theta}_0^i)$
$y_1^i \sim h_1(\cdot|x_1^i, \bar{\theta}_0^i)$
$\theta_1^i = \bar{\theta}_0^i$

estimate of the densities:

$$p_1^n(x_1, \theta_1|y_1) = \frac{\sum_{i=1}^n K_{h_n}^y(y_1 - y_1^i)\ K_{h_n}^\theta(\theta_1 - \theta_1^i)\ K_{h_n}^x(x_1 - x_1^i)}{\sum_{i=1}^n K_{h_n}^y(y_1 - y_1^i)}$$

$$p_1^n(\theta_1|y_1) = \frac{\sum_{i=1}^n K_{h_n}^y(y_1 - y_1^i)\ K_{h_n}^\theta(\theta_1 - \theta_1^i)}{\sum_{i=1}^n K_{h_n}^y(y_1 - y_1^i)}$$

$$p_1^n(x_1|y_1) = \frac{\sum_{i=1}^n K_{h_n}^y(y_1 - y_1^i)\ K_{h_n}^x(x_1 - x_1^i)}{\sum_{i=1}^n K_{h_n}^y(y_1 - y_1^i)}$$

For $t \geq 2$

generation of the trajectories: for $i = 1 \cdots n$

$(\bar{x}_{t-1}^i, \bar{\theta}_{t-1}^i) \sim p_{t-1}^n(x_{t-1}, \theta_{t-1}|y_{1:t-1})$
$x_t^i \sim f_t(\cdot|\bar{x}_{t-1}^i, \bar{\theta}_{t-1}^i)$
$y_t^i \sim h_t(\cdot|x_t^i, \bar{\theta}_{t-1}^i)$
$\theta_t^i = \bar{\theta}_{t-1}^i$

estimate of the densities:

$$p_t^n(x_t, \theta_t|y_{1:t}) = \frac{\sum_{i=1}^n K_{h_n}^y(y_t - y_t^i)\ K_{h_n}^\theta(\theta_t - \theta_t^i)\ K_{h_n}^x(x_t - x_t^i)}{\sum_{i=1}^n K_{h_n}^y(y_t - y_t^i)}$$

$$p_t^n(\theta_t|y_{1:t}) = \frac{\sum_{i=1}^n K_{h_n}^y(y_t - y_t^i)\ K_{h_n}^\theta(\theta_t - \theta_t^i)}{\sum_{i=1}^n K_{h_n}^y(y_t - y_t^i)}$$

$$p_t^n(x_t|y_{1:t}) = \frac{\sum_{i=1}^n K_{h_n}^y(y_t - y_t^i)\ K_{h_n}^x(x_t - x_t^i)}{\sum_{i=1}^n K_{h_n}^y(y_t - y_t^i)}$$

Table 3: The resampled convolution filter for Bayesian estimation.

In practice, the parameter prior law $\rho(\theta)$, the number of particles n, the kernels K and the associated bandwidth parameters h_n must by chosen by the user.

6.2 Theoretical study

The L^1-convergence of the R-CF filter to the optimal filter for a general state-space, at t fixed and as the number of particle n tends to infinity is proved in [7]. This result applied to the particular state-space model (11) proves that the particle filter presented in the previous section converges to the optimal filter:

Theorem 1 *Suppose that K is a positive Parzen-Rosenblatt kernel, and that*

- $y_t \mapsto p(y_t|y_{1:t-1})$ *is continuous and strictly positive (at least in the neighborhood of the actual observation), for all* $t \geq 0$,
- $y_t \mapsto p(y_t|x_t,\theta) \leq M_t$ *for some* $M_t < \infty$ *and for all* $t \geq 0$,
- $\lim_{n\to\infty} \frac{n h_n^{q+d+p}}{\log n} = \infty$ *and* $h_n^{2q} = O(n^{-\alpha})$ *with* $\alpha \in (0,1)$.

Then, for any t fixed

$$\lim_{n\to\infty} \iint |p_t^n(x_t,\theta_t|y_{1:t}) - p(x_t,\theta_t|y_{1:t})|\, dx_t\, d\theta_t = 0 \quad a.s.$$

The previous algorithm leads to the following estimate of the parameter

$$\hat{\theta}_t^n \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n \bar{\theta}_t^i \quad \text{with} \quad \bar{\theta}_t^i \sim p_t^n(\theta_t|y_{1:t})\,. \tag{12}$$

The convergence of this estimator is proved in Appendix B:

Theorem 2 *In addition to the assumptions of Theorem 1, suppose that the support of $p(\theta_t|y_{1:t})$ is included in a known compact set C then*

$$\lim_{n\to\infty} \hat{\theta}_t^n = \mathbb{E}[\theta_t|y_{1:t}] \quad a.s.$$

The previous theorems ensure that our estimates converge to the posterior law of the parameters or to $\mathbb{E}[\theta|y_{1:t}]$. These quantities are consistent Bayesian estimates of θ as $t \to \infty$. See Schervish [24] for complete a study of consistency in the Bayesian approach.

Theorem 2 is transposable to the state x_t. More precisely, let $\bar{x}_t^i$ generated from $p_t^n(x,\theta|y_{1:t})$ a consistent estimate of $\mathbb{E}[x_t|y_{1:t}]$ is then $\hat{x}_t^n = \frac{1}{n}\sum_{i=1}^n \bar{x}_t^i$. This estimate is used in the comparisons of part VII-B.

7 Simulated case studies

In a first example, with a short length time series, we compare the off-line techniques presented in Sections 3 and 4 with the on–line technique proposed in Section 6. In a second example we compare the convolution filter (R-CF) with the standard particle filter (PF) and with the extended Kalman filter (EKF).

7.1 Comparison of the approaches

Consider the following system proposed by Lo [25] with an unknown state variance:

$$\begin{aligned} x_{t+1} &= 1.1\, e^{-2\,x_t^2} - 1 + \theta\, w_t\,, \quad x_0 \sim \mathcal{N}(-0.5, 0.1^2)\,, \\ y_t &= x_t^3 + 0.1\, v_t \end{aligned}$$

where w_t, v_t are independent standard white Gaussian noises. The true value of the parameter is $\theta_* = 0.5$.

7.1.1 Least squares estimate

The objective is to minimize the approximate least mean squares criterion (8) given by the CF with n particles. Since $\hat{Q}_T^n$ is a random function, to avoid stochastic minimization we work on a fixed realization of random quantities as explained in section 5. More precisely, in this particular case, the random variables w_t and v_t are assumed to be independent of everything and of course of θ. The values of the noises generated to build the filter with a given value of θ, are thus valid for any other value of θ. Hence by always preserving the same values for the noises, the function $\hat{Q}_T^n$ is not random any more, it is then possible to minimize it with the traditional algorithms. This is done with a Matlab native function initialized with 0 and 2 as lower and upper bound respectively.

In order to characterize their impact, we varied the numbers n of particles employed and T of observations. The results are presented in Tab. 4. It gathered the maximum, the standard deviation and the average of the absolute errors of the estimates of the parameter on 500 different trajectories.

It arises from the results presented that the quality of the estimates improves when the number T of observations increases. However, it is not clearly the case when the number of particles n increases, whereas theoretically this should be the case. Several reasons could explain this phenomenon: the choice of the h_n is not optimal and could be badly adapted for large n, the CF can be too degraded for T large.

Nevertheless, the quality of the estimates remains correct throughout the various case studies and by comparison, the stochastic minimization techniques are inordinately more time consuming.

Nb n of particles	$T=20$ Max	S-Dev	Mean	$T=50$ Max	S-Dev	Mean
$n=20$	1.08	0.17	0.18	0.56	0.09	0.11
$n=50$	0.74	0.15	0.16	0.47	0.08	0.10
$n=100$	0.82	0.16	0.17	0.52	0.08	0.09
$n=200$	0.83	0.16	0.17	0.68	0.08	0.09
$n=500$	0.86	0.17	0.18	0.74	0.10	0.10
$n=1000$	0.88	0.18	0.19	0.73	0.11	0.10
$n=5000$	0.79	0.18	0.18	0.94	0.12	0.12

Table 4: Absolute errors on 500 trajectories with T=20, 50 (LS)

7.1.2 Maximum likelihood estimate

The objective is to maximize the approximate likelihood function (10) provided by the CF filter. This function is random, according to the same considerations as for the conditional least squares estimate, we use the technique of minimization with random quantities fixed, exposed in Section 5. The context of simulation is the same as for the conditional least squares. The algorithm of minimization, a Matlab native function is initialized with 0 and 2 as lower and upper bound for θ. Table 5 shows the maximum, the standard deviation and the mean of the absolute errors, of the estimates on 500 different trajectories, for $T=20$ and $T=50$ respectively.

Nb n of particles	$T=20$ Max	Std-Dev	Mean	$T=50$ Max	Std-Dev	Mean
$n=20$	1.50	0.18	0.21	1.02	0.16	0.23
$n=50$	0.86	0.14	0.17	0.76	0.13	0.20
$n=100$	0.69	0.14	0.17	0.70	0.14	0.21
$n=200$	0.81	0.13	0.17	0.71	0.14	0.23
$n=500$	0.78	0.15	0.18	0.75	0.15	0.27
$n=1000$	0.90	0.15	0.19	0.76	0.16	0.31
$n=5000$	0.84	0.18	0.22	1.03	0.19	0.40

Table 5: Absolute errors on 500 trajectories with T=20 and T=50 (MLE)

The performances are significantly lower than those of conditional least squares. More surprisingly, the performances are lower with the larger value of T, contrary to the theoretical predictions.

This probably comes from the fact that the likelihood estimate is less accurate than the estimate of $\mathbb{E}[y_t|y_{1:t-1}]$. Indeed, for a fixed number of observations, there is more variability in a function estimate than in an expectation estimate.

The least squares minimization technique thus seems preferable to the maximization likelihood approach.

7.1.3 R-CF based parameter estimate

To keep conditions comparable with the preceding cases, we suppose here a uniform prior law on the interval $[0, 2]$ for the parameter, i.e. $\rho(\theta) = \mathcal{U}[0, 2]$. Here $c_x = c_y = c_\theta = 1$.

This approach is different from the two other based on optimization, since the parameter is estimated throughout filtering. The filter is run along $T = 120$ time steps and the value retained for the estimate of θ is $\hat{\theta}^n_{120}$ given by (12).

Table 6, shows: the maximum, the standard deviation and the mean of the absolute errors on the estimates of 500 trajectories at time $T = 120$ for various number of particles.

Nb particles	Max	Std-Dev	Mean
$n = 20$	1.62	0.22	0.24
$n = 50$	1.18	0.16	0.18
$n = 100$	1.01	0.11	0.13
$n = 200$	0.92	0.11	0.12
$n = 500$	0.53	0.08	0.09
$n = 1000$	0.39	0.07	0.08
$n = 5000$	0.45	0.05	0.06

Table 6: Absolute errors on 500 estimations (R-CF Unknown Param.)

The Bayesian approach features better performances than all the approaches previously studied. Moreover, in accordance with the theoretical results, the error significantly drops down as n increases. Note that this Bayesian technique is performed on line contrary to previous methods. It is thus possible to get an online estimate of the filter without having to wait for an estimate of the parameter. Moreover this method is even faster than the two other approaches.

7.2 Bearings–only tracking

We compare the convolution filter (R-CF) with the standard bootstrap particle filter (BPF) and with the extended Kalman filter (EKF) applied to the classical problem of bearings-only tracking in the plane (see, e.g., [26]).

Consider a mobile (the target) with a rectilinear uniform motion (i.e., with constant bearing and speed) in the plane. This mobile is tracked by an observer with a given trajectory. The state vector $x_t = (p^1_t, p^2_t, v^1_t, v^2_t)^*$ represents the relative positions and velocities vector of the Cartesian coordinates for the difference between the tracked object and the observer:

$$x_t = x^{\text{tg}}_t - x^{\text{obs}}_t \,.$$

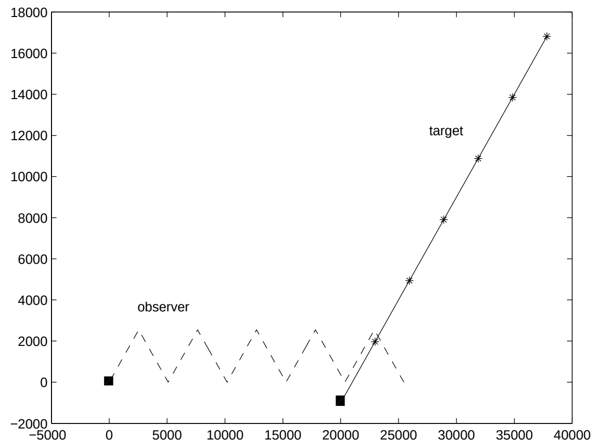

Figure 1: Simulation scenario. Total observation time 1 hour, sampling interval 4s. Initial relative distance 20025m, target speed 7m/s, observer speed 10m/s. Trajectories: target (plain line), maneuvering observer (dashed line), initial positions (black squares).

This state vector is solution of a linear noise–free system:

$$x_{t+1} = A_t\, x_t + B_t \tag{13}$$

where A_t and B_t are given.

The observations are a sequence of bearings corrupted by noise:

$$y_t = \tan^{-1}\Big(\frac{p_t^1}{p_t^2}\Big) + \sigma\, v_t$$

where v_t is a white Gaussian noise $\mathcal{N}(0,1)$ and $\sigma = 0.5$ degree. The simulation scenario is described in Fig. 1. The initial state law is $(p_0^1, p_0^2, v_t^1, v_t^2)^* \sim \mathcal{N}\big((23000, -3000, -10, 0)^*, \mathrm{diag}(5000^2, 5000^2, 10^2, 10^2)\big)$ while the true value is $(20000, -1000, -12, -2)^*$.

We perform 15 independent Monte Carlo runs of this scenario. In Figs. 2 to 6 we present the corresponding empirical position (the empirical estimated trajectory) and the corresponding empirical uncertainty ellipses (every 10 minutes). For R-CF and BPF we use 10000 particles. Calculation times and memory requirements are equivalent for R-CF and BPF.

This example is known to be unfavorable for EKF (see Fig. 2) but it does clearly show the advantage of our approach. Moreover, the standard particle filter requires the addition of an artificial noise in the state equation (13). The adjustment of the intensity of this noise is complicated, so it is a delicate process implementing the standard particle filter, see Figs. 3 and 4. Filter R-FC appears simpler and more robust in all the cases.

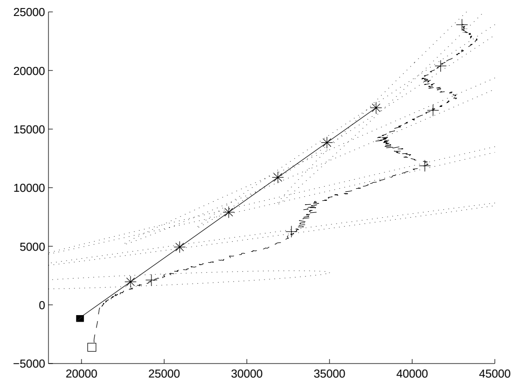

Figure 2: Extended Kalman filter. Plain line: true trajectory. Dashed line: empirical estimated trajectory after 15 Monte Carlo independent runs and the corresponding empirical uncertainty ellipses (every 10 minutes).

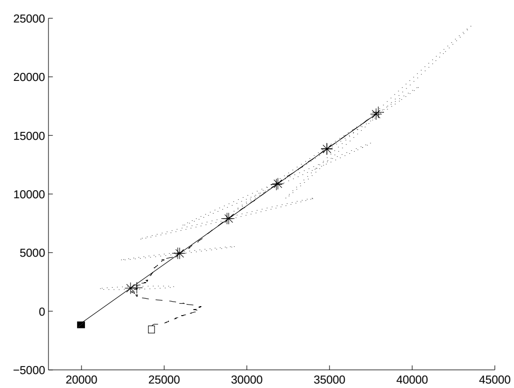

Figure 3: Bootstrap particle filter (BPF) with artificial Gaussian noise on (13): 0.025m/s standard deviation on the velocity components and 25m on the position components.

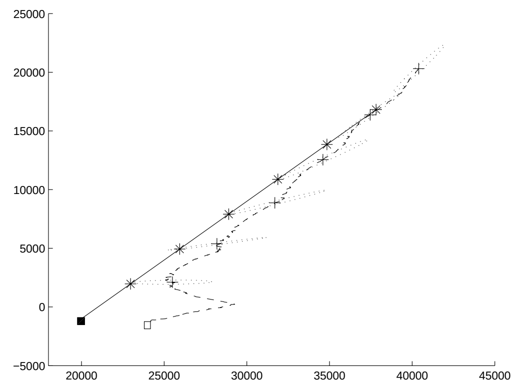

Figure 4: Bootstrap particle filter (BPF) with artificial Gaussian noise on (13): 0.05m/s standard deviation on the velocity components and 50m on the position components.

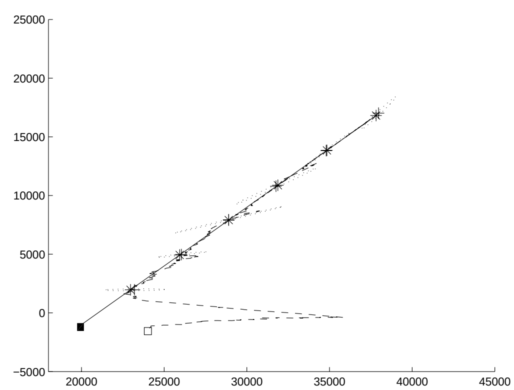

Figure 5: Resampled convolution filter (RCF): $c_x = 0.8$ and $c_y = 1$.

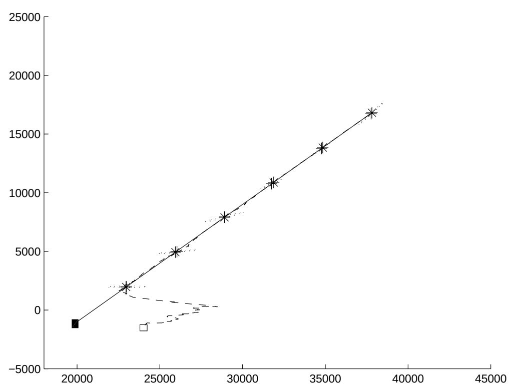

Figure 6: Resampled convolution filter (RCF): $c_x = 0.6$ and $c_y = 1$.

8 Conclusion and discussion

The different estimation approaches proposed show the large potential of the convolution filters.

The first approach, based on the maximization of the likelihood estimate and the minimization of the conditional least squares estimate, presents several drawbacks. From the practical point of view they require a high computation time and the choice of the number of observation T. From the theoretical point of view, their convergence is ensured under uniform convergence assumptions, which is difficult to verify for a given dynamic system. However, the convolution filters approach is a good alternative to the stochastic optimization, and can be used to perform the initialization of a Bayesian procedure.

The R-CF with unknown parameters approach introduced in the last section is perfectly suited for online estimation and their theoretical properties are clearly established without need of additional strong assumptions. Thus this last approach is interesting especially if the primary objective is the filtering in spite of uncertainties with the model.

A Kernel estimation

Definition 1 *A kernel $K : \mathbb{R}^d \mapsto \mathbb{R}$ is a bounded, positive, symmetric application such that $\int K(x)\,dx = 1$.*

We denote

$$K_{h_n}(x) \stackrel{\text{def}}{=} \frac{1}{h_n^d}\, K\Big(\frac{x}{h_n}\Big)\,.$$

$h_n > 0$ is the bandwidth parameter. The Gaussian kernel is $K(x) = \big(\frac{1}{\sqrt{2\pi}}\big)^d\, e^{-|x|^2/2}$.

Definition 2 *A Parzen-Rosenblatt kernel is a kernel such that* $\|x\|^d K(x) \to 0$ *as* $\|x\| \to \infty$.

Definition 3 *Let* $X_1 \cdots X_n$ *be i.i.d. random variables with common density* f. *The kernel estimator of* f, f_n, *associated with the kernel* K *is given by*

$$f_n(x) = \tfrac{1}{n\,h_n^d} \textstyle\sum_{i=1}^n K\big(\tfrac{x-X_i}{h_n}\big) = (K_{h_n} * \mu_n)(x)$$

for $x \in \mathbb{R}^d$; $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ *is the empirical measure associated with* $X_1 \cdots X_n$.

B Proof of theorem 2

Let

$$\Delta_n \stackrel{\text{def}}{=} \tfrac{1}{2} \textstyle\int |p_t^n(\theta_t|y_{1:t}) - p(\theta_t|y_{1:t})|\, d\theta_t \in [0,1]\,.$$

For notational convenience let

$$p_t^n = p_t^n(\theta_t|y_{1:t})\,, \quad p_t = p(\theta_t|y_{1:t})\,.$$

Consider a n-sample from the density p_t^n:

$$S_t^n \stackrel{\text{def}}{=} \{\bar\theta_t^1, \dots, \bar\theta_t^n\} \quad \text{with} \quad \bar\theta_t^i \sim p_t^n\,.$$

We show that there exists a subsample $\{\bar\theta_t^{i_1}, \dots, \bar\theta_t^{i_{M_n}}\} \subset S_t^n$, and a new sample $\{\dot\theta_t^{i_1}, \dots, \dot\theta_t^{i_{N_n}}\}$, which together can be considered as sampled from p_t. Such a device was used by [27] to study the robustness of kernel estimates.

Define

$$f_n \stackrel{\text{def}}{=} \tfrac{\min(p_t^n, p_t)}{1-\Delta_n}\,,\ g_n \stackrel{\text{def}}{=} \tfrac{p_t^n - \min(p_t^n, p_t)}{\Delta_n}\,,\ h_n \stackrel{\text{def}}{=} \tfrac{p_t - \min(p_t^n, p_t)}{\Delta_n}$$

they are density functions and

$$\begin{aligned} p_t^n &= \Delta_n\, g_n + (1-\Delta_n)\, f_n\,, \\ p_t &= \Delta_n\, h_n + (1-\Delta_n)\, f_n\,. \end{aligned}$$

This shows that each $\bar\theta_t^i$ sampled according to p_t^n is, with probability Δ_n, sampled from g_n.

Let

$$Z_i \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } \bar\theta_t^i \sim g_n\,, \\ 0 & \text{if } \bar\theta_t^i \sim f_n\,. \end{cases}$$

and

$$N_n \stackrel{\text{def}}{=} \textstyle\sum_{i=1}^n Z_i\,, \qquad\qquad M_n \stackrel{\text{def}}{=} n - N_n\,.$$

The Z_i's are Bernoulli variables with parameter Δ_n, so $N_n \sim \mathcal{B}(n, \Delta_n)$ is binomial. M_n is the number of $\bar\theta_t^i$'s sampled from f_n. Let $\{\bar\theta_t^{i_1}, \dots, \bar\theta_t^{i_{M_n}}\}$ be this subsample, $1 \le i_1 < \cdots < i_{M_n} \le n$. Let $I_M = \{i_1, \dots, i_{M_n}\}$ and $I_N = \{1, \dots, n\} \setminus I_{M_n}$.

Consider now the new following random sample:

$$\tilde{\theta}_t^i = \begin{cases} \bar{\theta}_t^i & \text{if } i \in I_M \\ \dot{\theta}_t^i & \text{with } \dot{\theta}_t^i \sim h_n,\ \text{if } i \in I_N \end{cases} \tag{14}$$

for $i = 1 \cdots n$. $\{\tilde{\theta}_t^1, \ldots, \tilde{\theta}_t^n\}$ can be considered as a virtual random sample from the unknown p_t which holds M_n common elements with S_t^n drawn from p_t^n. Let $\tilde{\vartheta}_t^n \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n \tilde{\theta}_t^i$ the associated virtual estimate of θ_t. Note that

$$|\hat{\theta}_t^n - \mathbb{E}[\theta_t|y_{1:t}]| \leq |\hat{\theta}_t^n - \tilde{\vartheta}_t^n| + |\tilde{\vartheta}_t^n - \mathbb{E}[\theta_t|y_{1:t}]| \tag{15}$$

As $\{\tilde{\theta}_t^1, \ldots, \tilde{\theta}_t^n\}$ are sampled from p_t, the strong law of large numbers ensures that $\lim_{n\to\infty} |\tilde{\vartheta}_t^n - \mathbb{E}[\theta_t|y_{1:t}]| = 0$ a.s. It remains to study the first term of the r.h.s. of (15). As $\{\tilde{\theta}_t^1, \ldots, \tilde{\theta}_t^n\}$ and $\{\bar{\theta}_t^1, \ldots, \bar{\theta}_t^n\}$ have M_n common elements, we have

$$|\hat{\theta}_t^n - \tilde{\vartheta}_t^n| = \tfrac{1}{n}| \textstyle\sum_{j=1}^{N_n} \bar{\theta}_t^{i_j} - \sum_{j=1}^{N_n} \tilde{\theta}_t^{i_j}| \leq \tfrac{2N_n}{n} \max_C |\theta|\,.$$

However $\frac{N_n}{n}$ is the empirical estimate of Δ_n, by Hoeffding's Inequality [28], for any Δ_n,

$$\mathbb{P}\Big(\Big|\frac{N_n}{n} - \Delta_n\Big| \geq \varepsilon\Big) \leq 2\exp\{-2n\varepsilon^2\}. \tag{16}$$

By Theorem 1 we have $\Delta_n \to 0$ a.s., then (16) implies that $\frac{N_n}{n} \to 0$ a.s. The proof is then completed.

References

[1] P. Del Moral, *Feynman-Kac formulae*. New York: Springer–Verlag, 2004.

[2] A. Doucet, N. de Freitas, and N. J. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. New York: Springer–Verlag, 2001.

[3] D. Crişan and A. Doucet, "A survey of convergence results on particle filtering for practitioners," *IEEE Transactions on Signal Processing*, vol. 50, no. 3, pp. 736–746, March 2002.

[4] A. Doucet and V. B. Tadić, "Parameter estimation in general state–space models using particle methods," *Annals of the Institute of Statistical Mathematics*, vol. 55, no. 2, pp. 409–422, 2003.

[5] J. Liu and M. West, "Combined parameter and state estimation in simulation-based filtering," in *Sequential Monte Carlo Methods in Practice*, ser. Statistics for Engineering and Information Science, A. Doucet, N. de Freitas, and N. J. Gordon, Eds. New York: Springer–Verlag, 2001, pp. 197–223.

[6] F. Cérou, F. Le Gland, and N. Newton, "Stochastic particle methods for linear tangent filtering equations," in *Optimal Control and Partial Differential Equations. In honour of professor Alain Bensoussan's 60th birthday*, J.-L. Menaldi, E. Rofman, and A. Sulem, Eds. Amsterdam: IOS Press, 2001, pp. 231–240.

[7] J.-P. Villa, V. Rossi, "Nonlinear filtering in discret time: A particle convolution approach," Biostatistic group of Montpellier, Technical Report 04-03, 2004, (available at http://vrossi.free.fr/recherche.html).

[8] L. Devroye and G. Lugosi, *Combinatorial Methods in Density Estimation*. Springer Verlag, 2001.

[9] H. Tong, *Nonlinear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press, 1990.

[10] L. Klimko and P. Nelson, "On conditional least squares estimation for stochastic processes," *The Annals of Statistics*, vol. 6, pp. 629–642, 1978.

[11] V. Rossi, "Filtrage non linéaire par noyaux de convolution. Application à un bioprocédé de dépollution," Ph.D. dissertation, ENSAM, 2004.

[12] G. Kitagawa, "Non–Gaussian state space modeling of nonstationary time series (with discussion)," *Journal of the American Statistical Association*, vol. 82, no. 400, pp. 1032–1063, Dec. 1987.

[13] ——, "Monte Carlo filter and smoother for non–Gaussian nonlinear state space models," *Journal of Computational and Graphical Statistics*, vol. 5, no. 1, pp. 1–25, 1996.

[14] C. Geyer, "Estimation and optimization of functions," in *Markov Chain Monte Carlo in practice*, W.R. Gilks, S. Richardson, D.J. Spiegelhalter Eds. London: Chapman & Hall, 1996, pp. 241–258.

[15] M. Hürzeler and H. Künsch, "Approximating and maximising the likelihood for a general state-space model," in *Sequential Monte Carlo Methods in Practice*, ser. Statistics for Engineering and Information Science, A. Doucet, N. de Freitas, and N. J. Gordon, Eds. New York: Springer–Verlag, 2001, ch. 12, pp. 159–175.

[16] G. Storvik, "Particle filters in state space models with the presence of unknown static parameters," *IEEE Transactions on Signal Processing*, 2002.

[17] T. Higuchi, "Monte Carlo filter using the genetic algorithm operators," *Journal of Statistical Computation and Simulation*, vol. 59, no. 1, pp. 1–23, 1997.

[18] W. R. Gilks and C. Berzuini, "Following a moving target — Monte Carlo inference for dynamic Bayesian models," *Journal of the Royal Statistical Society, Series B*, vol. 63, no. 1, pp. 127–146, 2001.

[19] D. S. Lee and N. K. K. Chia, "A particle algorithm for sequential Bayesian parameter estimation and model selection," *IEEE Transactions on Signal Processing*, vol. 50, pp. 326–336, 2002.

[20] M. West, "Approximating posterior distributions by mixtures," *Journal of the Royal Statistical Society, Series B*, vol. 55, no. 2, pp. 409–422, 1993.

[21] P. Del Moral and L. Miclo, "Branching and interacting particle systems approximations of Feynman–Kac formulae with applications to nonlinear filtering," in *Séminaire de Probabilités XXXIV*, ser. Lecture Notes in Mathematics, J. Azéma, M. Émery, M. Ledoux, and M. Yor, Eds. Berlin: Springer–Verlag, 2000, vol. 1729, pp. 1–145.

[22] F. Le Gland and N. Oudjane, "A robustification approach to stability and to uniform particle approximation of nonlinear filters: the example of pseudo-mixing signals," *Stochastic Processes and their Applications*, vol. 106, no. 2, pp. 279–316, 2003.

[23] P. Del Moral, J. Jacod, and P. Protter, "The Monte Carlo method for filtering with discrete–time observations," *Probability Theory and Related Fields*, vol. 120, no. 3, pp. 346–368, 2001.

[24] M. Schervish, *Theory of statistics*. Springer–Verlag, 1995.

[25] J. T. Lo, "Synthetic approach to optimal filtering," *IEEE Transactions on Neural Networks*, 1994.

[26] B. Ristic, M. S. Arulampalam, and N. J. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House, 2004.

[27] L. Devroye, *A Course on Density Estimation*, ser. Progress in Probability and Statistics. Boston: Birkhäuser, 1987, vol. 14.

[28] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, pp. 13–30, 1963.